Channel Polarization on q-ary Discrete Memoryless Channels by Arbitrary Kernels

Ryuhei Mori
Graduate School of Informatics
Kyoto University
Kyoto, 606–8501, Japan
Email: rmori@sys.i.kyoto-u.ac.jp

Toshiyuki Tanaka
Graduate School of Informatics
Kyoto University
Kyoto, 606–8501, Japan
Email: tt@i.kyoto-u.ac.jp

***Abstract*—A method of channel polarization, proposed by Arıkan, allows us to construct efficient capacity-achieving channel codes. In the original work, binary input discrete memoryless channels are considered. A special case of q-ary channel polarization is considered by Şaşoğlu, Telatar, and Arıkan. In this paper, we consider more general channel polarization on q-ary channels. We further show explicit constructions using Reed-Solomon codes, on which asymptotically fast channel polarization is induced.**

I. Introduction

Channel polarization, proposed by Arıkan, is a method of constructing capacity achieving codes with low encoding and decoding complexities [1]. Channel polarization can also be used to construct lossy source codes which achieve rate-distortion trade-off with low encoding and decoding complexities [2]. Arıkan and Telatar derived the rate of channel polarization [3]. In [4], a more detailed rate of channel polarization which includes coding rate is derived. In [1], channel polarization is based on a 2×2 matrix. Korada, Şaşoğlu, and Urbanke considered generalized polarization phenomenon which is based on an $\ell \times \ell$ matrix and derived the rate of the generalized channel polarization [5]. In [6], a special case of channel polarization on q-ary channels is considered. In this paper, we consider channel polarization on q-ary channels which is based on arbitrary mappings.

II. Preliminaries

Let $u_0^{\ell-1}$ and u_i^j denote a row vector $(u_0, \dots, u_{\ell-1})$ and its subvector $(u_i, \dots, u_j)$. Let $\mathcal{F}^c$ denote the complement of a set $\mathcal{F}$, and $|\mathcal{F}|$ denotes cardinality of $\mathcal{F}$. Let $\mathcal{X}$ and $\mathcal{Y}$ be an input alphabet and an output alphabet, respectively. In this paper, we assume that $\mathcal{X}$ is finite and that $\mathcal{Y}$ is at most countable. A discrete memoryless channel (DMC) W is defined as a conditional probability distribution $W(y \mid x)$ over $\mathcal{Y}$ where $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. We write $W : \mathcal{X} \to \mathcal{Y}$ to mean a DMC W with an input alphabet $\mathcal{X}$ and an output alphabet $\mathcal{Y}$. Let q be the cardinality of $\mathcal{X}$. In this paper, the base of the logarithm is q unless otherwise stated.

Definition 1: The symmetric capacity of q-ary input channel $W : \mathcal{X} \to \mathcal{Y}$ is defined as

$$I(W) := \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} \frac{1}{q} W(y \mid x) \log \frac{W(y \mid x)}{\frac{1}{q}\sum_{x'\in\mathcal{X}} W(y \mid x')}.$$

Note that $I(W) \in [0, 1]$.

Definition 2: Let $\mathcal{D}_x := \{y \in \mathcal{Y} \mid W(y \mid x) > W(y \mid x'), \forall x' \in \mathcal{X}, x' \neq x\}$. The error probability of the maximum-likelihood estimation of the input x on the basis of the output y of the channel W is defined as

$$P_e(W) := \frac{1}{q} \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{D}_x^c} W(y \mid x).$$

Definition 3: The Bhattacharyya parameter of W is defined as

$$Z(W) := \frac{1}{q(q-1)} \sum_{\substack{x\in\mathcal{X}, x'\in\mathcal{X},\\ x\neq x'}} Z_{x,x'}(W)$$

where the Bhattacharyya parameter of W between x and x' is defined as

$$Z_{x,x'}(W) := \sum_{y\in\mathcal{Y}} \sqrt{W(y \mid x) W(y \mid x')}.$$

The symmetric capacity $I(W)$, the error probability $P_e(W)$, and the Bhattacharyya parameter $Z(W)$ are interrelated as in the following lemmas.

Lemma 4:

$$P_e(W) \leq (q-1) Z(W).$$

Lemma 5: [6]

$$\begin{aligned} I(W) &\geq \log \frac{q}{1 + (q-1)Z(W)} \\ I(W) &\leq \log(q/2) + (\log 2)\sqrt{1 - Z(W)^2} \\ I(W) &\leq 2(q-1)(\log \mathrm{e})\sqrt{1 - Z(W)^2}. \end{aligned}$$

Definition 6: The maximum and the minimum of the Bhattacharyya parameters between two symbols are defined as

$$\begin{aligned} Z_{\max}(W) &:= \max_{x\in\mathcal{X}, x'\in\mathcal{X}, x\neq x'} Z_{x,x'}(W) \\ Z_{\min}(W) &:= \min_{x\in\mathcal{X}, x'\in\mathcal{X}} Z_{x,x'}(W). \end{aligned}$$

Let $\sigma : \mathcal{X} \to \mathcal{X}$ be a permutation. Let σ^i denote the ith power of σ. The average Bhattacharyya parameter of W between x and x' with respect to σ is defined as the average of

$Z_{z,z'}(W)$ over the subset $\{(z,z') = (\sigma^i(x), \sigma^i(x')) \in \mathcal{X}^2 \mid i = 0,1,\dots,q!-1\}$ as

$$Z^{\sigma}_{x,x'}(W) := \frac{1}{q!}\sum_{i=0}^{q!-1} Z_{\sigma^i(x),\sigma^i(x')}(W).$$

III. Channel polarization on q-ary DMC induced by non-linear kernel

We consider a channel transform using a one-to-one onto mapping $g : \mathcal{X}^\ell \to \mathcal{X}^\ell$, which is called a kernel. In the previous works [1], [5], it is assumed that $q = 2$ and that g is linear. In [6], $\mathcal{X}$ is arbitrary but g is restricted. In this paper, $\mathcal{X}$ and g are arbitrary.

Definition 7: Let $W : \mathcal{X} \to \mathcal{Y}$ be a DMC. Let $W^\ell : \mathcal{X}^\ell \to \mathcal{Y}^\ell$, $W^{(i)} : \mathcal{X} \to \mathcal{Y}^\ell \times \mathcal{X}^{i-1}$, and $W^{(i)}_{u_0^{i-1}} : \mathcal{X} \to \mathcal{Y}^\ell$ be defined as DMCs with transition probabilities

$$W^\ell(y_0^{\ell-1} \mid x_0^{\ell-1}) := \prod_{i=0}^{\ell-1} W(y_i \mid x_i)$$
$$W^{(i)}(y_0^{\ell-1}, u_0^{i-1} \mid u_i) := \frac{1}{q^{\ell-1}} \sum_{u_{i+1}^{\ell-1}} W^\ell(y_0^{\ell-1} \mid g(u_0^{\ell-1}))$$
$$W^{(i)}_{u_0^{i-1}}(y_0^{\ell-1} \mid u_i) := \frac{1}{q^{\ell-i-1}} \sum_{u_{i+1}^{\ell-1}} W^\ell(y_0^{\ell-1} \mid g(u_0^{\ell-1})).$$

Definition 8: Let $\{B_i\}_{i=0,1,\dots}$ be independent random variables such that $B_i = k$ with probability $\frac{1}{\ell}$, for each $k = 0,\dots,\ell-1$.

In probabilistic channel transform $W \to W^{(B_i)}$, expectation of the symmetric capacity is invariant due to the chain rule for mutual information. The following lemma is a consequence of the martingale convergence theorem.

Lemma 9: There exists a random variable I_∞ such that $I(W^{(B_0)\cdots(B_n)})$ converges to I_∞ almost surely as $n \to \infty$.

When $q = 2$ and $g(u_0^1) = (u_0 + u_1, u_1)$, Arıkan showed that $P(I_\infty \in \{0,1\}) = 1$ [1]. This result is called channel polarization phenomenon since subchannels polarize to noiseless channels and pure noise channels. Korada, Şaşoğlu, and Urbanke consider channel polarization phenomenon when $q = 2$ and g is linear [5].

From Lemma 5, $I(W)$ is close to 0 and 1 when $Z(W)$ is close to 1 and 0, respectively. Hence, it would be sufficient to prove channel polarization if one can show that $Z(W^{(B_1)\cdots(B_n)})$ converges to $Z_\infty \in \{0,1\}$ almost surely. Here we instead show a weaker version of the above property in the following lemma and its corollary.

Lemma 10: Let $\{\mathcal{Y}_n\}_{n\in\mathbb{N}}$ be a sequence of discrete sets. Let $\{W_n : \mathcal{X} \to \mathcal{Y}_n\}_{n\in\mathbb{N}}$ be a sequence of q-ary DMCs. Let σ and τ be permutations on $\mathcal{X}$. Let

$$W'_n(y_1, y_2 \mid x) = W_n(y_1 \mid \sigma(x)) W_n(y_2 \mid \tau(x))$$

where $W_n : \mathcal{X} \to \mathcal{Y}_n$, $W'_n : \mathcal{X} \to \mathcal{Y}_n^2$. Assume $\lim_{n\to\infty} I(W'_n) - I(W_n) = 0$. Then, for any $\delta \in (0, 1/2)$, there exists m such that $Z^{\tau\sigma^{-1}}_{x,x'}(W_n) \notin (\delta, 1-\delta)$ for any $x \in \mathcal{X}$, $x' \in \mathcal{X}$ and $n \ge m$.

Proof: Let Z, Y_1 and Y_2 be random variables which take values on $\mathcal{X}$, $\mathcal{Y}_n$ and $\mathcal{Y}_n$, respectively, and jointly obey the distribution

$$P_n(Z = z, Y_1 = y_1, Y_2 = y_2) = \frac{1}{q} W_n(y_1 \mid \sigma(z)) W_n(y_2 \mid \tau(z)).$$

Since $I(W'_n) = I(Z; Y_1, Y_2)$ and $I(W_n) = I(Z; Y_1)$,

$$I(Z; Y_1, Y_2) - I(Z; Y_1) = I(Z; Y_2 \mid Y_1)$$

tends to 0 by the assumption. Since the mutual information is lower bounded by the cut-off rate, one obtains

$$\begin{aligned}
I(Z; Y_2 \mid Y_1) &\ge -\log \sum_{y_1\in\mathcal{Y}_n, y_2\in\mathcal{Y}_n} P_n(Y_1 = y_1) \\
&\quad \times \left[\sum_{z\in\mathcal{X}} P_n(Z = z \mid Y_1 = y_1) \times \sqrt{P_n(Y_2 = y_2 \mid Z = z, Y_1 = y_1)}\right]^2 \\
&= -\log \sum_{y_1\in\mathcal{Y}_n, z\in\mathcal{X}, x\in\mathcal{X}} P_n(Y_1 = y_1) P_n(Z = z \mid Y_1 = y_1) \\
&\quad \times P_n(Z = x \mid Y_1 = y_1) Z_{\tau(z),\tau(x)}(W_n) \\
&= -\log \sum_{y_1\in\mathcal{Y}_n, z\in\mathcal{X}, x\in\mathcal{X}} q_n(y_1, z, x) Z_{\tau(\sigma^{-1}(z)),\tau(\sigma^{-1}(x))}(W_n)
\end{aligned}$$

where

$$q_n(y_1, z, x) := P_n(Y_1 = y_1) \times P_n(Z = \sigma^{-1}(z) \mid Y_1 = y_1) P_n(Z = \sigma^{-1}(x) \mid Y_1 = y_1).$$

Since

$$\begin{aligned}
\sum_{y_1\in\mathcal{Y}} q_n(y_1, z, x) &= \sum_{y_1\in\mathcal{Y}} P_n(Y_1 = y_1) \\
&\quad \times \left(\sqrt{P_n(Z = \sigma^{-1}(z) \mid Y_1 = y_1) P_n(Z = \sigma^{-1}(x) \mid Y_1 = y_1)}\right)^2 \\
&\ge \left(\sum_{y_1\in\mathcal{Y}} P_n(Y_1 = y_1) \times \sqrt{P_n(Z = \sigma^{-1}(z) \mid Y_1 = y_1) P_n(Z = \sigma^{-1}(x) \mid Y_1 = y_1)}\right)^2 \\
&= \frac{1}{q^2} Z_{z,x}(W_n)^2
\end{aligned}$$

it holds

$$I(Z; Y_2 \mid Y_1) \ge -\log \left[1 - \frac{1}{q^2} \sum_{z\in\mathcal{X}, x\in\mathcal{X}} Z_{z,x}(W_n)^2 (1 - Z_{\tau(\sigma^{-1}(z)),\tau(\sigma^{-1}(x))}(W_n))\right].$$

The convergence of $I(Z; Y_2 \mid Y_1)$ to 0 implies that

$$Z_{z,x}(W_n)^2 (1 - Z_{\tau(\sigma^{-1}(z)),\tau(\sigma^{-1}(x))}(W_n))$$

converges to 0 for any $(z, x) \in \mathcal{X}^2$. It consequently implies that for any $\delta \in (0, 1/2)$, there exists m such that

$Z_{x,x'}^{\tau\sigma^{-1}}(W_n) \notin (\delta, 1-\delta)$ for any $x \in \mathcal{X}$, $x' \in \mathcal{X}$ and $n \geq m$. ∎

Using Lemma 10, one can obtain a partial result of the channel polarization as follows.

Corollary 11: Assume that there exists $u_0^{\ell-2} \in \mathcal{X}^{\ell-1}$, $(i,j) \in \{0,1,\dots,\ell-1\}^2$ and permutations σ and τ on $\mathcal{X}$ such that i-th element of $g(u_0^{\ell-1})$ and j-th element of $g(u_0^{\ell-1})$ are $\sigma(u_{\ell-1})$ and $\tau(u_{\ell-1})$, respectively, and such that for any $v_0^{\ell-2} \neq u_0^{\ell-2} \in \mathcal{X}^{\ell-1}$ there exists $m \in \{0,1,\dots,\ell-1\}$ and a permutation μ on $\mathcal{X}$ such that m-th element of $g(v_0^{\ell-1})$ is $\mu(v_{\ell-1})$. Then, for almost every sequence $b_1,\dots,b_n,\dots$ of $0,\dots,\ell-1$, and for any $\delta \in (0,1/2)$, there exists m such that $Z_{x,x'}^{\tau\sigma^{-1}}(W^{(b_1)\cdots(b_n)}) \notin (\delta, 1-\delta)$ for any $x \in \mathcal{X}$, $x' \in \mathcal{X}$ and $n \geq m$.

Proof: Since $I(W^{(B_1)\cdots(B_n)})$ converges to I_∞ almost surely, $|I(W^{(B_1)\cdots(B_n)(\ell-1)}) - I(W^{(B_1)\cdots(B_n)})|$ has to converge to 0 almost surely. Let $U_0^{\ell-1}$ and $Y_0^{\ell-1}$ denote random variables ranging over $\mathcal{X}^\ell$ and $\mathcal{Y}^\ell$, and obeying the distribution

$$P(U_0^i = u_0^{\ell-1}, Y_0^{\ell-1} = y_0^{\ell-1}) = \frac{1}{q} W^{(\ell-1)}(y_0^{\ell-1}, u_0^{\ell-2} \mid u_{\ell-1}).$$

Then, it holds

$$\begin{aligned} I(W^{(\ell-1)}) &= I(Y_0^{\ell-1}, U_0^{\ell-2}; U_{\ell-1}) \\ &= I(Y_0^{\ell-1}; U_{\ell-1} \mid U_0^{\ell-2}) \\ &= \sum_{u_0^{\ell-2}} \frac{1}{q^{\ell-1}} I(Y_0^{\ell-1}; U_{\ell-1} \mid U_0^{\ell-2} = u_0^{\ell-2}). \end{aligned}$$

From the assumption, $I(Y_0^{\ell-1}; U_{\ell-1} \mid U_0^{\ell-2} = u_0^{\ell-2}) \geq I(W)$ for all $u_0^{\ell-2} \in \mathcal{X}^{\ell-1}$. Hence, $I(W^{(B_1)\cdots(B_n)}) - I(W^{(B_1)\cdots(B_n)})$ has to converge to 0 almost surely. By applying Lemma 10, one obtains the result. ∎

When $q = 2$, since $Z(W) = Z_{0,1}(W)$, this corollary immediately implies the channel polarization phenomenon, although it is not sufficient for general $q \neq 2$. Note that in this derivation one does not use extra conditions e.g., symmetricity of DMC, linearity of a kernel.

If a kernel is linear, a more detailed condition is obtained.

Definition 12: Assume $(\mathcal{X}, +, \cdot)$ be a commutative ring. A kernel $g : \mathcal{X}^\ell \to \mathcal{X}^\ell$ is said to be linear if $g(ax + bz) = ag(x) + bg(z)$ for all $a \in \mathcal{X}$, $b \in \mathcal{X}$, $x \in \mathcal{X}^\ell$, and $z \in \mathcal{X}^\ell$.

If g is linear, g can be represented by a square matrix G such that $g(u_0^{\ell-1}) = u_0^{\ell-1}G$. Let $U_0^{\ell-1}$, $X_0^{\ell-1}$ and $Y_0^{\ell-1}$ denote random variables taking values on $\mathcal{X}^\ell$, $\mathcal{X}^\ell$ and $\mathcal{Y}^\ell$, respectively, and obeying distribution

$$\begin{aligned} P(U_0^{\ell-1} &= u_0^{\ell-1}, X_0^{\ell-1} = x_0^{\ell-1}, Y_0^{\ell-1} = y_0^{\ell-1}) \\ &= \frac{1}{2^\ell} W^\ell\left(y_0^{\ell-1} \mid u_0^{\ell-1}G\right) \mathbb{I}\{x_0^{\ell-1}V = u_0^{\ell-1}\} \end{aligned}$$

where V denotes an $\ell \times \ell$ full-rank upper triangle matrix. There exists a one-to-one correspondence between X_0^i and U_0^i for all $i \in \{0,\dots,\ell-1\}$. Hence, statistical properties of $W^{(i)}$ are invariant under an operation $G \to VG$. Further, a permutation of columns of G does not change statistical properties of $W^{(i)}$ either. Since any full-rank matrix can be decomposed to the form VLP where V, L, and P are upper triangle, lower triangle, and permutation matrices, without loss of generality we assume that G is a lower triangle matrix and that $G_{kk} = 1$ where $k \in \{0,\dots,\ell-1\}$ is the largest number such that the number of non-zero elements in k-th row of G is greater than 1, and where G_{ij} denotes (i,j) element of G.

Theorem 13: Assume that $\mathcal{X}$ is a field of prime cardinality, and that linear kernel G is not diagonal. Then, $P(I_\infty \in \{0,1\}) = 1$.

Proof: It holds

$$\begin{aligned} W^{(k)}(y_0^{\ell-1}, u_0^{k-1} \mid u_k) &= \frac{1}{q^{\ell-1}} \prod_{j=k+1}^{\ell-1} \left(\sum_{x \in \mathcal{X}} W(y_j \mid x) \right) \\ &\times \prod_{j \in S_0} W(y_j \mid x_j) \prod_{j \in S_1} W(y_j \mid G_{kj}u_k + x_j) \end{aligned}$$

where $S_0 := \{j \in \{0,\dots,\ell-1\} \mid G_{kj} = 0\}$, $S_1 := \{j \in \{0,\dots,\ell-1\} \mid G_{kj} \neq 0\}$, and x_j is j-th element of $(u_0^{k-1}, 0_k^{\ell-1})G$ where $0_k^{\ell-1}$ is all-zero vector of length $\ell - k$. Let $m \in \{0,\dots,k-1\}$ be such that $G_{km} \neq 0$. Since each u_0^{k-1} occurs with positive probability $1/q^k$, we can apply Lemma 10 with $\sigma(x) = x$ and $\tau(x) = G_{km}x + z$ for arbitrary $z \in \mathcal{X}$. Hence, for sufficiently large n, $Z_{x,x'}^{\mu}(W^{(B_1)\cdots(B_n)})$ is close to 0 or 1 almost surely where $\mu(x) = G_{km}^i x + z$ for all $i \in \{0,\dots,q-2\}$ and $z \in \mathcal{X}$. Since q is a prime, when $\mu_0(z) = z + x' - x$ for $x \neq x'$, $Z_{x,x'}^{\mu_0}(W^{(B_1)\cdots(B_n)})$ is close to 0 or 1 if and only if $Z(W^{(B_1)\cdots(B_n)})$ is close to 0 or 1, respectively. ∎

This result is a simple generalization of the special case considered by Şaşoğlu, Telatar, and Arıkan [6]. For a prime power q and a finite field $\mathcal{X}$, we show a sufficient condition for channel polarization in the following corollary.

Corollary 14: Assume that $\mathcal{X}$ is a field and that a linear kernel G is not diagonal. If there exists $j \in \{0,\dots,k-1\}$ such that G_{kj} is a primitive element. Then, $P(I_\infty \in \{0,1\}) = 1$.

Proof: By applying Lemma 10, one sees that for almost every sequence $b_1,\dots,b_n,\dots$ of $0,\dots,\ell-1$, and for any $\delta \in (0,1/2)$, there exists m such that $Z_{x,x'}^{\sigma}(W^{(b_1)\cdots(b_n)}) \notin (\delta, 1-\delta)$ for any $x \in \mathcal{X}$, $x' \in \mathcal{X}$ and $n \geq m$ where $\sigma(x) = G_{kj}x + z$ for arbitrary $z \in \mathcal{X}$. It suffices to show that for any $x \in \mathcal{X}$ and $x' \in \mathcal{X}$, $x \neq x'$ $Z_{x,x'}(W^{(B_1)\cdots(B_n)})$ is close to 1 if and only if $Z(W^{(B_1)\cdots(B_n)})$ is close to 1. When $Z_{x,x'}(W^{(B_1)\cdots(B_n)})$ is close to 1, $Z_{0,G_{kj}(x'-x)}(W^{(B_1)\cdots(B_n)})$ is close to 1. Hence, $Z_{0,G_{kj}^i(x'-x)}(W^{(B_1)\cdots(B_n)})$ is close to 1 for any $i \in \{0,\dots,q-2\}$. Since G_{kj} is a primitive element, $Z_{0,x}(W^{(B_1)\cdots(B_n)})$ is close to 1 for any $x \in \mathcal{X}$. It completes the proof. ∎

In [7], it is shown that the channel polarization phenomenon occurs by using a random kernel in which G_{kj} is chosen uniformly from nonzero elements. Corollary 14 says that a deterministic primitive element G_{kj} is sufficient for the channel polarization phenomenon.

IV. Speed of Polarization

Arıkan and Telatar showed the speed of polarization [3]. Korada, Şaşoğlu, and Urbanke generalized it to any binary linear kernels [5].

Proposition 15: Let $\{\hat{X}_n \in (0,1)\}_{n\in\mathbb{N}}$ be a random process satisfying the following properties.

1) $\hat{X}_n$ converges to $\hat{X}_\infty$ almost surely.
2) $\hat{X}_{n+1} \le \hat{c}\hat{X}_n^{\hat{D}_n}$ where $\{\hat{D}_n \ge 1\}_{n\in\mathbb{N}}$ are independent and identically distributed random variables, and $\hat{c}$ is a constant.

Then,

$$\lim_{n\to\infty} P(\hat{X}_n < 2^{-2^{\beta n}}) = P(\hat{X}_\infty = 0)$$

for $\beta < \mathbb{E}[\log_2 \hat{D}_1]$ where $\mathbb{E}[\cdot]$ denotes an expectation. Similarly, let $\{\check{X}_n \in (0,1)\}_{n\in\mathbb{N}}$ be a random process satisfying the following properties.

1) $\check{X}_n$ converges to $\check{X}_\infty$ almost surely.
2) $\check{X}_{n+1} \ge \check{c}\check{X}_n^{\check{D}_n}$ where $\{\check{D}_n \ge 1\}_{n\in\mathbb{N}}$ are independent and identically distributed random variables, and $\check{c}$ is a constant.

Then,

$$\lim_{n\to\infty} P(\check{X}_n < 2^{-2^{\beta n}}) = 0$$

for $\beta > \mathbb{E}[\log_2 \check{D}_1]$.
Note that the above proposition can straightforwardly be extended to include the rate dependence [4].

In order to apply Proposition 15 to $Z_{\max}(W^{(B_1)\cdots(B_n)})$ and $Z_{\min}(W^{(B_1)\cdots(B_n)})$ as $\hat{X}_n$ and $\check{X}_n$, respectively, the second conditions have to be proven. In the argument of [5], partial distance of a kernel corresponds to the random variables $\hat{D}_n$ and $\check{D}_n$ in Proposition 15.

Definition 16: Partial distance of a kernel $g : \mathcal{X}^\ell \to \mathcal{X}^\ell$ is defined as

$$\begin{aligned}&D^{(i)}_{x,x'}(u_0^{i-1})\\ &\quad := \min_{v_{i+1}^{\ell-1}, w_{i+1}^{\ell-1}} d(g(u_0^{i-1}, x, v_{i+1}^{\ell-1}), g(u_0^{i-1}, x', w_{i+1}^{\ell-1}))\end{aligned}$$

where $d(a,b)$ denotes the Hamming distance between $a \in \mathcal{X}^\ell$ and $b \in \mathcal{X}^\ell$.
We also use the following quantities.

$$\begin{aligned} D^{(i)}_{x,x'} &:= \min_{u_0^{i-1}} D^{(i)}_{x,x'}(u_0^{i-1})\\ D^{(i)}_{\max} &:= \max_{x\in\mathcal{X}, x'\in\mathcal{X}} D^{(i)}_{x,x'}\\ D^{(i)}_{\min} &:= \min_{\substack{x\in\mathcal{X}, x'\in\mathcal{X}\\ x\ne x'}} D^{(i)}_{x,x'}.\end{aligned}$$

When g is linear, $D^{(i)}_{x,x'}(u_0^{i-1})$ does not depend on x, x' or u_0^{i-1}, in which case we will use the notation $D^{(i)}$ instead of $D^{(i)}_{x,x'}(u_0^{i-1})$.

From Lemma 21 in the appendix, the following lemma is obtained.

Lemma 17: For $i \in \{0, \dots, \ell-1\}$,

$$\frac{1}{q^{2\ell-2-i}} Z_{\min}(W)^{D^{(i)}_{x,x'}} \le Z_{x,x'}(W_\ell^{(i)}) \le q^{\ell-1-i} Z_{\max}(W)^{D^{(i)}_{x,x'}}$$

Corollary 18: For $i \in \{0, \dots, \ell-1\}$,

$$\begin{aligned} Z_{\max}(W^{(i)}) &\le q^{\ell-1-i} Z_{\max}(W)^{D^{(i)}_{\min}}\\ \frac{1}{q^{2\ell-2-i}} Z_{\min}(W)^{D^{(i)}_{\max}} &\le Z_{\min}(W^{(i)}).\end{aligned}$$

From Proposition 15 and Corollary 18, the following theorem is obtained.

Theorem 19: Assume $P(I_\infty(W) \in \{0,1\}) = 1$. It holds

$$\lim_{n\to\infty} P(Z(W^{(B_1)\dots(B_n)}) < 2^{-\ell^{\beta n}}) = I(W)$$

for $\beta < (1/\ell)\sum_i \log_\ell D^{(i)}_{\min}$.

When $Z_{\min}(W) > 0$,

$$\lim_{n\to\infty} P(Z(W^{(B_1)\dots(B_n)}) < 2^{-\ell^{\beta n}}) = 0$$

for $\beta > (1/\ell)\sum_i \log_\ell D^{(i)}_{\max}$.

When g is a linear kernel represented by a square matrix G, $(1/\ell)\sum_i \log_\ell D^{(i)}$ is called the exponent of G [5].

Example 20: Assume that $\mathcal{X}$ is a field and that $\alpha \in \mathcal{X}$ is a primitive element. For a non-zero element $\gamma \in \mathcal{X}$, let

$$G_{\mathrm{RS}}(q) = \begin{bmatrix} 1 & 1 & \dots & 1 & 1 & 0\\ \alpha^{(q-2)(q-2)} & \alpha^{(q-3)(q-2)} & \dots & \alpha^{q-2} & 1 & 0\\ \alpha^{(q-2)(q-3)} & \alpha^{(q-3)(q-3)} & \dots & \alpha^{q-3} & 1 & 0\\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots\\ \alpha^{q-2} & \alpha^{q-3} & \dots & \alpha & 1 & 0\\ 1 & 1 & \dots & 1 & 1 & \gamma \end{bmatrix}.$$

Since $G_{\mathrm{RS}}(2) = \begin{bmatrix} 1 & 0\\ 1 & 1\end{bmatrix}$, $G_{\mathrm{RS}}(q)$ can be regarded as a generalization of Arıkan's original matrix. The relation between binary polar codes and binary Reed-Muller codes [1] also holds for q-ary polar codes using $G_{\mathrm{RS}}(q)$ and q-ary Reed-Muller codes. From Theorem 13, the channel polarization phenomenon occurs on $G_{\mathrm{RS}}(q)$ for any $\gamma \ne 0$ when q is a prime. When γ is a primitive element, from Corollary 14, the channel polarization phenomenon occurs on $G_{\mathrm{RS}}(q)$ for any prime power q. We call $G_{\mathrm{RS}}(q)$ the Reed-Solomon kernel since the submatrix which consists of i-th row to $(q-1)$-th row of $G_{\mathrm{RS}}(q)$ is a generator matrix of a generalized Reed-Solomon code, which is a maximum distance separable code i.e., $D^{(i)} = i+1$. Hence, the exponent of $G_{\mathrm{RS}}(q)$ is $\frac{1}{\ell}\sum_i \log_\ell(i+1)$ where $\ell = q$. Since

$$\frac{1}{\ell}\sum_{i=0}^{\ell-1} \log_\ell(i+1) \ge \frac{1}{\ell\log_{\mathrm{e}}\ell}\int_1^\ell \log_{\mathrm{e}} x \mathrm{d}x = 1 - \frac{\ell-1}{\ell\log_{\mathrm{e}}\ell}$$

the exponent of the Reed-Solomon kernel tends to 1 as $\ell = q$ tends to infinity. When $q = 2^2$, the exponent of the Reed-Solomon kernel is $\log_{\mathrm{e}} 24/(4\log_{\mathrm{e}} 4) \approx 0.57312$. In Arıkan's original work, the exponent of the 2×2 matrix is 0.5 [3]. In [5], Korada, Şaşoğlu, and Urbanke showed that by using large kernels, the exponent can be improved, and found a matrix of size 16 whose exponent is about 0.51828. The above-mentioned Reed-Solomon kernel with $q = 2^2$ is reasonably small and simple but has a larger exponent than binary linear

kernels of small size. This demonstrates the usefulness of considering q-ary rather than binary channels. For q-ary DMC where q is not a prime, it can be decomposed to subchannels of input sizes of prime numbers [7] by using the method of multilevel coding [8]. The above example shows that when q is a power of a prime, without the decomposition of q-ary DMC, asymptotically better coding scheme can be constructed by using q-ary polar codes with $G_{\mathrm{RS}}(q)$.

V. Conclusion

The channel polarization phenomenon on q-ary channels has been considered. We give several sufficient conditions on kernels under which the channel polarization phenomenon occurs. We also show an explicit construction with a q-ary linear kernel $G_{\mathrm{RS}}(q)$ for q being a power of a prime. The exponent of $G_{\mathrm{RS}}(q)$ is $\log_{\mathrm{e}}(q!)/(q\log_{\mathrm{e}} q)$ which is larger than the exponent of binary matrices of small size even if $q=4$. Our discussion includes channel polarization on non-linear kernels as well. It is known that non-linear binary codes may have a larger minimum distance than linear binary codes, e.g. the Nordstrom-Robinson codes [9]. This implies possibility that there exists a non-linear kernel with a larger exponent than any linear kernel of the same size.

Appendix

Lemma 21:

$$\frac{1}{q^{2(\ell-1-i)}} Z_{\min}(W)^{D^{(i)}_{x,x'}(u_0^{i-1})} \le Z_{x,x'}(W^{(i)}_{u_0^{i-1}}) \le q^{\ell-1-i} Z_{\max}(W)^{D^{(i)}_{x,x'}(u_0^{i-1})}$$

Proof: For the second inequality, one has

$$\begin{aligned}
Z_{x,x'}(W^{(i)}_{u_0^{i-1}}) &= \sum_{y_0^{\ell-1}} \sqrt{W^{(i)}_{u_0^{i-1}}(y_0^{\ell-1}\mid x) W^{(i)}_{u_0^{i-1}}(y_0^{\ell-1}\mid x')} \\
&= q^i \sum_{y_0^{\ell-1}} \sqrt{W^{(i)}(y_0^{\ell-1}, u_0^{i-1}\mid x) W^{(i)}(y_0^{\ell-1}, u_0^{i-1}\mid x')} \\
&= \frac{1}{q^{\ell-1-i}} \sum_{y_0^{\ell-1}} \Bigg(\sum_{v_{i+1}^{\ell-1}, w_{i+1}^{\ell-1}} W^{\ell}(y_0^{\ell-1}\mid u_0^{i-1}, x, v_{i+1}^{\ell-1}) W^{\ell}(y_0^{\ell-1}\mid u_0^{i-1}, x', w_{i+1}^{\ell-1}) \Bigg)^{\frac{1}{2}} \\
&\le \frac{1}{q^{\ell-1-i}} \sum_{y_0^{\ell-1}} \sum_{v_{i+1}^{\ell-1}, w_{i+1}^{\ell-1}} \sqrt{W^{\ell}(y_0^{\ell-1}\mid u_0^{i-1}, x, v_{i+1}^{\ell-1}) W^{\ell}(y_0^{\ell-1}\mid u_0^{i-1}, x', w_{i+1}^{\ell-1})} \\
&\le \frac{1}{q^{\ell-1-i}} \sum_{v_{i+1}^{\ell-1}, w_{i+1}^{\ell-1}} Z_{\max}(W)^{D^{(i)}_{x,x'}(u_0^{i-1})} \\
&= q^{\ell-1-i} Z_{\max}(W)^{D^{(i)}_{x,x'}(u_0^{i-1})}.
\end{aligned}$$

The first inequality is obtained as follows.

$$\begin{aligned}
Z_{x,x'}(W^{(i)}_{u_0^{i-1}}) &= \sum_{y_0^{\ell-1}} \sqrt{W^{(i)}_{u_0^{i-1}}(y_0^{\ell-1}\mid x) W^{(i)}_{u_0^{i-1}}(y_0^{\ell-1}\mid x')} \\
&= q^i \sum_{y_0^{\ell-1}} \sqrt{W^{(i)}(y_0^{\ell-1}, u_0^{i-1}\mid x) W^{(i)}(y_0^{\ell-1}, u_0^{i-1}\mid x')} \\
&= \sum_{y_0^{\ell-1}} \Bigg(\sum_{v_{i+1}^{\ell-1}, w_{i+1}^{\ell-1}} \frac{1}{q^{2(\ell-1-i)}} \times W^{\ell}(y_0^{\ell-1}\mid u_0^{i-1}, x, v_{i+1}^{\ell-1}) W^{\ell}(y_0^{\ell-1}\mid u_0^{i-1}, x', w_{i+1}^{\ell-1}) \Bigg)^{\frac{1}{2}} \\
&\ge \sum_{y_0^{\ell-1}} \sum_{v_{i+1}^{\ell-1}, w_{i+1}^{\ell-1}} \frac{1}{q^{2(\ell-1-i)}} \times \sqrt{W^{\ell}(y_0^{\ell-1}\mid u_0^{i-1}, x, v_{i+1}^{\ell-1}) W^{\ell}(y_0^{\ell-1}\mid u_0^{i-1}, x', w_{i+1}^{\ell-1})} \\
&\ge \frac{1}{q^{2(\ell-1-i)}} Z_{\min}(W)^{D^{(i)}_{x,x'}(u_0^{i-1})}.
\end{aligned}$$

■

Acknowledgment

TT acknowledges support of the Grant-in-Aid for Scientific Research on Priority Areas (No. 18079010), MEXT, Japan.

References

[1] E. Arıkan, “Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels,” *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3051–3073, July 2009.

[2] S. Korada and R. Urbanke, “Polar codes are optimal for lossy source coding,” 2009. [Online]. Available: http://arxiv.org/abs/0903.0307

[3] E. Arıkan and E. Telatar, “On the rate of channel polarization,” in *Proc. 2009 IEEE International Symposium on Information Theory*, June 28-July 3 2009, pp. 1493–1495.

[4] T. Tanaka and R. Mori, “Refined rate of channel polarization,” 2010. [Online]. Available: http://arxiv.org/abs/1001.2067

[5] S. Korada, E. Şaşoğlu, and R. Urbanke, “Polar codes: characterization of exponent, bounds, and constructions,” 2009. [Online]. Available: http://arxiv.org/abs/0901.0536

[6] E. Şaşoğlu, E. Telatar, and E. Arıkan, “Polarization for arbitrary discrete memoryless channels,” 2009. [Online]. Available: http://arxiv.org/abs/0908.0302

[7] E. Sasoglu, E. Telatar, and E. Arikan, “Polarization for arbitrary discrete memoryless channels,” in *Proc. 2009 IEEE Information Theory Workshop, Taormina, Italy*, 11–16 Oct. 2009, pp. 144–148.

[8] H. Imai and S. Hirakawa, “A new multilevel coding method using error-correcting codes,” *Information Theory, IEEE Transactions on*, vol. 23, no. 3, pp. 371–377, may 1977.

[9] F. MacWilliams and N. Sloane, *The Theory of Error-Correcting Codes*. North-Holland Amsterdam, 1977.